DOCUMENT RESUME

ED 036 871 | 24 | CG 005 170

AUTHOR	Guilford, J.P.
TITLE	Comments on Professor Bloom's Paper Entitled "Toward a Theory of Testing Which Includes Measurement-Evaluation-Assessment". Center for the study of Evaluation of Instructional Programs Occasional Report No. 12.
INSTITUTION	California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO	BR-6-1646
PUB DATE	Jun 68
CONTRACT	OEC-4-6-061646-1909
NOTE	11p.
EDRS PRICE	EDRS Price MF-$0.25 HC-$0.65
DESCRIPTORS	*Critical Reading, *Education, Educational Testing, *Evaluation, Learning, *Measurement, Objectives, *Testing

ABSTRACT

This commentary analyzes Professor Bloom's definitions of kinds of evaluation and the needs for evaluation in education. In a discussion of the nature of tests of cognition, memory, and production and evaluation abilities, Professor Guilford stresses the need for concern with acquisition of specific items of information and with general intellectual skills for dealing with that information. (Author)

ED036871

PA 6-1641
OE BR

Occasional Report No. 12

PA-24

COMMENTS ON PROFESSOR BLOOM'S PAPER ENTITLED "TOWARD A THEORY OF TESTING WHICH INCLUDES MEASUREMENT-EVALUATION-ASSESSMENT"

J. P. Guilford

CG005170

Center FOR THE
Study of
Evaluation
OF INSTRUCTIONAL
PROGRAMS

University of California, Los Angeles, June 1968

CO-DIRECTORS

Merlin C. Wittrock Erick L. Lindman

ASSOCIATE DIRECTORS

Marvin C. Alkin Frank Massey, Jr. C. Robert Pace

The CENTER FOR THE STUDY OF EVALUATION OF INSTRUCTIONAL PROGRAMS is engaged in research that will yield new ideas and new tools capable of analyzing and evaluating instruction. Staff members are creating new ways to evaluate content of curricula, methods of teaching and the multiple effects of both on students. The CENTER is unique because of its access to Southern California's elementary, secondary and higher schools of diverse socio-economic levels and cultural backgrounds. Three major aspects of the program are

Instructional Variables - Research in this area will be concerned with identifying and evaluating the effects of instructional variables, and with the development of conceptual models, learning theory and theory of instruction. The research involves the experimental study of the effects of differences in instruction as they may interact with individual differences among students.

Contextual Variables - Research in this area will be concerned with measuring and evaluating differences in community and school environments and the interactions of both with instructional programs. It will also involve evaluating variations in student and teacher characteristics and administrative organization.

Criterion Measures - Research in this field is concerned with creating a new conceptualization of evaluation of instruction and in developing new instruments to evaluate knowledge acquired in school by measuring observable changes in cognitive, affective and physiological behavior. It will also involve evaluating the cost-effectiveness of instructional programs.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED036871

COMMENTS ON PROFESSOR BLOOM'S PAPER ENTITLED "TOWARD A THEORY OF TESTING WHICH INCLUDES MEASUREMENT-EVALUATION-ASSESSMENT"

J. P. Guilford

The research and development reported herein was performed pursuant to a contract with the United States Department of Health, Education, and Welfare, Office of Education under the provisions of the Cooperative Research Program.

(Each paper printed as a Technical Report, Occasional Report, or Working Paper of the Center for the Study of Evaluation of Instructional Programs (UCLA) is reviewed prior to acceptance for publication. Standard journal review procedures are followed, which include submission of comments to the Director by professional reviewers. This insures that the technical competence of the papers will be maintained at the high level set by the Center.)

CSEIP Occasional Report No. 12, June 1968
University of California, Los Angeles

COMMENTS ON PROFESSOR BLOOM'S PAPER ENTITLED "TOWARD A THEORY OF TESTING WHICH INCLUDES MEASUREMENT-EVALUATION-ASSESSMENT"

Professor Bloom's paper reflects considerable thought to problems of measurement in education. In saying this, I am using the term "measurement" in my familiar broad sense and not in the limited sense in which Bloom chooses to apply it, namely, to those concerned with basic psychological traits. The paper considers the broad range of places at which measurements are needed in education and the reasons for those needs. Types of techniques are mentioned and where they apply. Varieties of reliability, validity, and norms are discussed as well as the purposes that they serve.

The paper is not so much about a synthesis of methods of measurement as it is a systematic survey, with comparisons and assignment of roles. Since psychological tests of basic traits and achievement examinations have had common use for many years and the assessment procedures (in the narrow sense) have not, it might be said that he is making a plea for the addition of those techniques. He more clearly makes a plea for more attention to the environment of the student. This means quantitative descriptions of environments on the one hand, and taking environmental conditions somehow into account in measuring traits of individuals, on the other. Just how the latter is to be achieved is not made clear. There is also a plea for more theory, which includes psychological theory, in

connection with the question of what is being measured. He contrasts the apparent wealth of theory on the part of those who deal in assessment procedures and the apparent poverty of theory on the part of the testers, or what he calls the "measurement" approach. The contrast actually seems exaggerated, however, for some testers have been very much concerned about theory, and they possess and they use more rigorous methods for testing their theories.

There are many excellent points made in the paper to which one can agree. Again we see a warning against the misuse of testing. All of us probably know instances in which some very bad decisions have been made, based on rigid interpretations of IQ's and other scores. Those making such decisions are functioning like technicians rather than as sophisticated, professional psychologists. The wrong use of tests can do much harm, but I should hesitate to go as far as Bloom, when he speaks of the potential of tests for destroying mankind as being equal to that of atomic energy. I sometimes wonder, however, what effect the widespread use of answer sheet tests may have had on our population. Our extensive experiences in the Aptitudes Research Project at the University of Southern California has demonstrated many times over that one cannot measure abilities for productive thinking, divergent or convergent, with one or two possible exceptions, by means of answer sheet tests. There are even a few cognition abilities (where cognition is defined in the restricted sense of the structure of intellect) that require completion items, not multiple-choice.

It is easy to agree with Bloom that "evaluation" or measurement of achievement in education should be in terms of the objectives

that have been set up for education in an area of instruction. This principle is often given lip service, but not so often observed in practice. A corollary to this principle, a very important one, is that the objectives should be so clearly spelled out that examination items can be written for each one of them. The objectives should often be as specific as the items themselves. Another corollary, which Bloom mentions, is that where objectives differ, examinations should differ. This calls into question the overemphasis on national testing programs and national norms.

It is very true, as Bloom says, that the kinds of tests that we apply influence the learner in his learning and the teacher in his teaching. They both work toward the end that the student shall do well in the tests. Tests also determine certain educational values, which, in turn, determine social values. For years, the IQ has helped to set educational goals. We have tried to see to it that each student shall perform educationally at a level consistent with his IQ. Now the IQ test is weighted heavily with cognition abilities (cognition in the structure-of-intellect sense), which represent only one-fifth of all known or expected intellectual abilities. The student can achieve in this respect just by understanding and absorbing information; there is little or no premium in also learning how to use that information in productive thinking.

I cannot agree with Bloom when he says that the psychological testers (whom he calls measurement people) assume that individuals who take their tests have had equal environmental opportunities. There may have been a day when developers of tests of abilities thought that what they wanted to measure is entirely determined by

heredity. Although test theorists, following Spearman, have recognized that every test score has an error component, I cannot recall anyone saying that he regarded that error to be completely contributed by the environment of the individual. I think it is safe to say that most testers regard any individual's score as being a function of both the person's heredity and his environment, and the true component is not necessarily attributed entirely to his heredity. The individual's score, allowing for its error component, tells us how the person stood on a certain scale at a certain time, without telling us how he got that way. It would take information from different sources to tell us how he got that way.

What I have just said applies more strictly to cognition tests. If I may refer to the structure of intellect again as a frame of reference, I can point out some exceptions. Cognition tests tell us how much information of a certain kind the examinee has in his possession. We do not know how or when he obtained it. In tests of memory abilities, however, we must ensure that examinees have had equal opportunity to learn the information on which we are going to test them. We therefore apply experimental controls, exposing them for a constant period of time to the same stimulus material. As a further control, in order to minimize or exclude cognition variance, the selected information to which they are exposed is made so easy to cognize that on a cognition test of it, they would all make perfect scores. For the measurement of production abilities, divergent or convergent, and evaluative abilities, we also apply the latter control, staying well within the range of common experience for all individuals tested. We do not always succeed in this, but we try. Factor analysis tells us when we have not succeeded.

As I read the section regarding evaluation, I had the impression that the interest in gain scores is overemphasized at the expense of status scores. The measurement of change offers numerous problems, which Chester Harris is well prepared to tell us about. There are problems of scaling so that numerical differences on one part of a scale are equivalent to those on other parts of the same scale. Some kind of absolute scaling seems called for. Furthermore, reliabilities of gain scores, in the form of differences between status scores, are notoriously unreliable. Rarely would they be sufficiently reliable for the purpose of individual measurement and there would be little use for norms. They would be sufficiently reliable for research on groups.

I agree with Bloom's concern about gaining information concerning the student's environment, past and present. In general, psychologists have paid too little attention to human environments. We need very much to know what relevant features and variables should be made known and should be measured in relation to behavioral variables. But I am puzzled by the insistence that information about the environment should somehow enter into the measurement of psychological and educational variables. Nor are we told how this can or should be done. Information regarding the environment is often very useful in understanding an individual's scores, but why should we combine that information with measures of the individual? I hope that I have not misinterpreted Bloom's intention.

A survey of available techniques for quantitative descriptions of students is useful, but I think that Bloom would agree that this is not the best place to start in planning a comprehensive program

in education. The first question to ask is for what aspects of personal development are the schools responsible? In this connection, what information do we need or want about individuals? No technique should be used just because it is available. If there are aspects of development for which no techniques of evaluation exist, we should see that those techniques are developed. There are other considerations. Is the method efficient and economical? Is what it has to tell us worth the effort? Will it arouse student or parental resistance? Will someone use the information that the method provides, and use it wisely?

There is one aspect of measurement in the form of evaluation that Bloom touched upon but which deserves greater emphasis. This is the aspect of continual feedback information, which measurement provides to the student as well as to the teacher, administrator, and counselor. The teacher should want to know how well the educational objectives are being fulfilled in the class that he teaches. Where are the weak spots and what kind of weakness exists? The serious student, like all motivated humans, wants to know, "How well am I doing?" He may be satisfied to know the answer in terms of a general quantity, such as a score or a grade. What he may not know, and we as psychologists do know, from the laws of learning, is that he would profit even more by having specific feedback information. It would be wise to arrange matters so that there is prompt and specific feedback to the student at every step of his learning.

At one time I knew a professor of chemistry who proposed a procedure and a kind of device that I am sure would be a big step forward in education. It would provide for individual testing of

students during a lecture. After making a particular point in lecturing, the teacher would give the class a multiple-choice test item on that point. Each student would press one of several buttons on his chair, which has a wired connection with a device on the lecturer's table. On a screen visible to teacher and students would flash the correct answer, also the number of correct answers. In the device on the table each student's score would be cumulated.

We are approaching this kind of operation, of course, in computerized learning. But I am sure that you will agree that we are far from realizing all the potential that our electronic age has made possible. My plea is that we give much more time to evaluation than we do and that it also be made an integral part of the teaching process, taking advantage of the best learning principles that we know.

As to broader aspects of educational evaluation, I should like to propose a general approach to which I have given some thought, without coming to any concrete procedural decisions. So far as the intellectual aspects of school learning are concerned, we have a two-fold obligation to the student: (a) to see that he acquires the desirable items of specific information, and (b) to see that he develops general, intellectual skills for dealing with that information. Together, these aspects make up what should be included in the individual's total intelligence. The first of these is now fairly well evaluated in terms of standard achievement examinations. The second is measured by tests of intellectual abilities. By this I do not mean that we be content with present IQ tests and academic-aptitude tests, for they do not go nearly so far as they should and

are limited to one or two scores. We are learning a great deal concerning the numerous unique intellectual abilities, which can be regarded as being equivalent to the generalized intellectual skills just mentioned. I do not contend that all of them would be of interest to the educator at all age levels or for all school subjects, but I am sure that many of them should be of serious educational interest in relevant places; and their periodic measurement should provide valuable information about the development of individuals. A program that involves such assessments should include sophisticated personnel who know how to use such information.